Supplemental Material for Joint Embedding of 3D Scan and CAD Objects

Manuel Dahnert[1] Angela Dai[1] Leonidas Guibas[2,3] Matthias Nießner[1]
[1]Technical University of Munich [2]Stanford University [3]Facebook AI Research

A. Additional Quantitative Studies

We provide several additional quantitative experiments evaluating robustness against rotation, as well as the performance of the object segmentation and completion from our stacked hourglass model.

Robustness to rotations To achiever robustness to potential rotations for input scan queries, we train our method with rotation augmentation around the up axis. During training, we rotate the initial partial and cluttered scan object as well as the positive and negative CAD model with the same random rotation. During evaluation of CAD model retrieval, we embed CAD models into the embedding space using 12 uniform rotations for each CAD model; for an input scan query, we then find the closest CAD embedding. We train for 160k iterations and a triplet margin of 0.1. Table 2 shows the results of CAD model retrieval while testing on randomly rotated scan object inputs. With this rotation augmentation, we can achieve performance on par with the case of canonically oriented objects while testing on arbitrarily rotated scan inputs: 0.42 in instance average retrieval accuracy and 0.16 in instance average ranking quality.

Method	**IoU**
[A] SGPN [2]	0.10
[B] Segmentation(Ours)	**0.36**
[C] Segmentation(Ours) + 3D-EPN [1]	0.48
[D] Segmentation(Ours) + Completion(Ours)	**0.53**

Table 1: Evaluation (IoU) of our segmentation and completion to SGPN [2] and 3D-EPN [1], respectively.

What is the performance of the segmentation and completion? In Table 1, we evaluate the performance of the first and second hourglass with Intersection over Union (IoU) between the predicted and ground truth binary occupancy grid. We compare our model against SGPN [2], a point cloud based segmentation method, and 3D-EPN [1], a voxel-based object completion network. For evaluation, we then convert all outputs to occupancy grids to compute the final IoU scores.

Additionally, we evaluate our stacked hourglass model, replacing our completion encoder-decoder with the model of 3D-EPN, trained end-to-end to learn a joint scan-CAD embedding. Tables 3, 4, and 5 show that our model achieves notably better performance in embedding space confusion as well as CAD model retrieval and ranking than the version using 3D-EPN.

References

[1] Angela Dai, Charles Ruizhongtai Qi, and Matthias Nießner. Shape completion using 3d-encoder-predictor cnns and shape synthesis. In *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 2017. 1, 2

[2] Weiyue Wang, Ronald Yu, Qiangui Huang, and Ulrich Neumann. Sgpn: Similarity group proposal network for 3d point cloud instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2569–2578, 2018. 1

Method	trash bin	bathtub	bed	bookshelf	cabinet	chair	display	file	sofa	table	other	**class avg**	**inst avg**
Ours with rotations	0.35	0.47	0.45	0.18	0.30	0.56	0.45	0.08	0.41	0.41	0.26	0.36	0.42

Table 2: Evaluation of CAD model retrieval by retrieval accuracy on our Scan-CAD Object Similarity benchmark.

Method	trash bin	bathtub	bed	bookshelf	cabinet	chair	display	file	sofa	table	**class avg**	**inst (k=10)**	**inst (k=50)**
Ours (no seg, no cmpl)	0.14	0.13	0.23	0.11	0.07	0.15	0.14	0.28	0.19	0.18	0.16	0.14	0.22
Ours (no cmpl)	0.24	0.32	0.26	0.28	0.13	0.21	0.44	0.24	0.19	0.25	0.24	0.21	0.31
Ours (no seg)	**0.50**	0.53	0.52	**0.51**	0.48	0.44	0.51	0.53	0.47	**0.50**	0.49	0.48	0.49
Ours (no triplet)	0.51	0.48	0.45	0.22	0.42	0.34	0.25	**0.50**	0.28	0.38	0.36	0.34	0.42
Ours (3D-EPN [1] for cmpl)	0.45	**0.51**	0.49	0.46	**0.50**	0.33	0.40	0.53	0.47	0.45	0.43	0.42	0.47
Ours (w/o end-to-end)	0.42	0.46	0.46	0.35	0.42	0.35	0.33	0.51	0.34	0.41	0.39	0.37	0.44
Ours	0.51	0.52	**0.50**	**0.51**	0.51	**0.48**	**0.50**	0.55	**0.51**	0.49	**0.50**	**0.49**	**0.50**

Table 3: Evaluation of the joint scan-CAD embedding space. We compare our learned scan-CAD feature space to those constructed from features computed through both handcrafted and learned shape descriptors. We evaluate the confusion between scan and CAD, where 0.5 reflects a perfect confusion.

Method	trash bin	bathtub	bed	bookshelf	cabinet	chair	display	file	sofa	table	other	**class avg**	**inst avg**
Ours (no seg, no cmpl)	0.06	0.00	0.15	0.04	0.00	0.47	0.30	0.00	0.20	0.13	0.04	0.13	0.23
Ours (no cmpl)	0.13	0.07	0.15	0.12	0.04	0.37	0.38	0.00	0.15	0.26	0.09	0.16	0.24
Ours (no seg)	0.14	0.07	0.24	0.13	0.15	0.40	0.32	0.17	0.15	0.21	0.13	0.19	0.26
Ours (no triplet)	0.03	0.13	0.39	0.04	0.11	0.07	0.08	0.00	0.13	0.09	0.04	0.10	0.08
Ours (3D-EPN [1] for cmpl)	0.41	0.33	0.42	**0.21**	0.19	0.49	0.40	0.08	0.20	0.30	0.31	0.30	0.37
Ours (w/o end-to-end)	0.42	0.27	**0.48**	0.07	0.15	0.42	0.27	**0.25**	**0.35**	0.21	0.32	0.29	0.32
Ours	**0.50**	**0.60**	0.42	0.19	**0.26**	**0.55**	**0.45**	**0.25**	0.33	**0.32**	**0.43**	**0.39**	**0.43**

Table 4: Evaluation of CAD model retrieval by top-1 retrieval accuracy on the test split of our Scan-CAD Object Similarity benchmark.

Method	trash bin	bathtub	bed	bookshelf	cabinet	chair	display	file	sofa	table	other	**class avg**	**inst avg**
Ours (no seg, no cmpl)	0.05	0.00	0.08	0.03	0.01	0.17	0.14	0.00	0.10	0.04	0.04	0.06	0.10
Ours (no cmpl)	0.08	0.00	0.06	0.04	0.02	0.15	0.12	0.06	0.06	**0.11**	0.05	0.07	0.10
Ours (no seg)	0.08	0.06	0.12	0.08	0.09	0.14	0.09	0.06	0.07	0.07	0.04	0.08	0.10
Ours (no triplet)	0.01	0.06	0.13	0.03	0.04	0.03	0.02	0.06	0.04	0.04	0.05	0.05	0.04
Ours (3D-EPN [1] for cmpl)	0.17	0.18	**0.19**	**0.09**	**0.12**	0.17	**0.15**	0.06	0.10	**0.11**	0.09	0.13	0.14
Ours (w/o end-to-end)	0.14	0.18	0.12	0.04	0.06	0.18	0.14	0.13	**0.16**	0.08	**0.12**	0.12	0.13
Ours	**0.29**	**0.24**	**0.19**	0.08	**0.12**	**0.19**	0.14	**0.19**	0.15	0.10	0.09	**0.16**	**0.16**

Table 5: Evaluation of CAD model retrieval by ranking quality on the test split of our Scan-CAD Object Similarity benchmark.